Exercise 1: Risk Minimizers for Generalized L2-Loss

Consider the regression learning setting, i.e., $\mathcal{Y} = \mathbb{R}$, and assume that your loss function of interest is $L(y, f(\mathbf{x})) = \big(m(y) - m(f(\mathbf{x}))\big)^2$, where $m : \mathbb{R} \to \mathbb{R}$ is a continuous strictly monotone function.

Disclaimer: In the following we always assume that $\mathsf{Var}(m(Y))$ exists.

(a) Consider the hypothesis space of constant models $\mathcal{H} = \{f : \mathcal{X} \to \mathbb{R} \,|\, f(\mathbf{x}) = \boldsymbol{\theta} \; \forall \mathbf{x} \in \mathcal{X}\}$, where $\mathcal{X}$ is the feature space. Show that

$$\hat{f}(\mathbf{x}) = m^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} m(y^{(i)})\right)$$

is the optimal constant model for the loss function above, where m^{-1} is the inverse function of m.

Hint: We can obtain several different notions of a mean value by using a specific function m, e.g., the arithmetic mean by $m(x) = x$, the harmonic mean by $m(x) = 1/x$ (if $x > 0$) or the geometric mean by $m(x) = \log(x)$ (if $x > 0$).

(b) Verify that the risk of the optimal constant model is $\mathcal{R}_L\left(\hat{f}\right) = \left(1 + \frac{1}{n}\right)\mathsf{Var}(m(y))$.

(c) Derive that the risk minimizer (Bayes optimal model) f^* is given by $f^*(\mathbf{x}) = m^{-1}\left(\mathbb{E}_{y|x}\left[m(y) \mid \mathbf{x}\right]\right)$.

(d) What is the optimal constant model in terms of the (theoretical) risk for the loss above and what is its risk?

(e) Recall the decomposition of the Bayes regret into the estimation and the approximation error. Show that the former is $\frac{1}{n}\mathsf{Var}(m(y))$, while the latter is $\mathsf{Var}\Big(\mathbb{E}_{y|x}\left[m(y) \mid \mathbf{x}\right]\Big)$ for the optimal constant model $\hat{f}(\mathbf{x})$ if the hypothesis space $\mathcal{H}$ consists of the constant models.

Hint: Use the law of total variance, which states that $\mathsf{Var}(Y) = \mathbb{E}_X\left[\mathsf{Var}(Y \mid X)\right] + \mathsf{Var}(\mathbb{E}_{Y|X}\left[Y \mid X\right])$, where the conditional variance is defined as $\mathsf{Var}_{Y|X}(Y \mid X) = \mathbb{E}_{Y|X}\left[\left(Y - \mathbb{E}_{Y|X}(Y \mid X)\right)^2 \mid X\right]$.